Adjusted Bayesian inference for selected parameters

Daniel Yekutieli

June 22, 2010

Abstract

We address the problem of providing inference for parameters selected after viewing the data from a Bayesian perspective. A frequentist solution to this problem is constructing False Coverage-statement Rate adjusted confidence intervals for the subset of selected parameters. We illustrate the limitations of the frequentist solution. We argue that if the parameter is elicited a non-informative prior, or if it is a "fixed" effect that is generated before selection is applied, then it is necessary to adjust the Bayesian inference for selection. Our main contribution is a Bayesian framework for providing inference for selected parameters, based on the observation that from a Bayesian perspective providing inference for a selected parameter is a truncation problem. Our second contribution is the introduction of Bayesian FDR controlling methodology, that generalizes existing Bayesian FDR methods to the case of non-dichotomous parameters. We illustrate our results by applying them to simulated data and data from a microarray experiment.

1 Introduction

The multiplicity problem is often identified in the statistical literature with the problem of selective and simultaneous inference. Benjamini and Yekutieli (2005) argue that the problem of selective inference and the simultaneity problem are two distinct problems encountered when trying to provide statistical inference for multiple parameters. Simultaneity refers to the need to provide inferences that apply to all the parameters, e.g. marginal confidence intervals that cover all the parameters with probability 0.95. A solution to this problem is Family Wise Error Rate adjusted inference. Selective inference refers to inference that is provided for parameters specified after viewing the data. The topic of this paper is Bayesian selective inference. We begin by describing a frequentist solution to the problem of selective inference, discussing selective inference in Genomic

association studies, and reviewing several aspects of Bayesian analysis that are relevant to our work.

1.1 Control over the false coverage-statement rate

Soric (1989) asserted that the goal of many scientific experiments is to discover non-zero effects, made the important observation that it is mainly the discoveries that are reported and included into science, and warned that unless the proportion of false discoveries in the set of declared discoveries is kept small there is danger that a large part of science is untrue.

Benjamini and Hochberg (1995), hereafter BH, considered the problem of testing m null hypotheses $H_1 \cdots H_m$, of which m_0 are true null hypotheses. They referred to the rejection of a null hypothesis a discovery and the rejection of a true null hypothesis a false discovery. To limit the occurrence of false discoveries when testing multiple null hypotheses BH introduced the False Discovery Rate $FDR = E\{V/\max(R,1)\}$, where R is the number of discoveries and V is the number false discoveries, and introduced the BH multiple testing procedure that controls the FDR at a nominal level q.

Benjamini and Yekutieli (2005) generalized the Benjamini and Hochberg testing framework. In their parameter selection framework there are m parameters $\theta_1 \cdots \theta_m$, with corresponding estimators $T_1 \cdots T_m$, and the goal is to construct valid confidence intervals (CIs) for the subset of parameters selected by a given selection rule $\mathcal{S}(t_1 \cdots t_m) \subseteq \{1 \cdots m\}$. They showed that CIs constructed for selected parameters no longer ensure nominal coverage probability, and suggested the False Coverage-statement Rate (FCR) as the appropriate criterion to capture the error for CIs constructed for selected parameters. The FCR is also defined $E\{V/\max(R,1)\}$, however R is the number of CIs constructed and V is the number of non-covering CIs. Benjamini and Yekutieli (2005) introduced a method of ensuring $FCR \leq q$ for independent $T_1 \cdots T_m$ and any selection criterion: construct marginal $1 - R \cdot q/m$ CIs for each of the R selected parameters. In cases where each θ_i can be associated with a null value θ_i^0 and the selection criteria are multiple testing procedures that test $\theta_i = \theta_i^0$ vs. $\theta_i \neq \theta_i^0$, Benjamini and Yekutieli (2005) showed that the level q BH procedure can be expressed as the least conservative multiple testing procedure that ensures that all level q FCR adjusted CI for θ_i, for which the null hypothesis is rejected, will not cover the respective θ_i^0. Furthermore, they show that for independent $T_1 \cdots T_m$ if all $\theta_i \neq \theta_i^0$ then applying the level q BH procedure to select the

parameters and declaring each selected θ_i greater than θ_i^0 if $T_i > \theta_i^0$ and smaller than θ_i^0 if $T_i < \theta_i^0$ controls the directional FDR (expected proportion of selected parameters assigned the wrong sign) at level $q/2$.

Example 1.1 Throughout the paper we use the following simulated example to illustrate the discussion. The simulation includes 10^5 iid samples of (θ_i, Y_i). To generate θ_i, we first sample λ_i from $\{10, 1\}$ with probabilities 0.90 and 0.10, and then draw θ_i from the absolute valued exponential density, $\pi_1(\theta_i|\lambda_i) = \lambda_i \cdot \exp(-\lambda_i \cdot |\theta_i|)/2$; $Y_i = \theta_i + \epsilon_i$, with ϵ_i independent $N(0, 1)$.

The selection rule is the level $q = 0.2$ BH procedure applied to the two sided p-values $p_i = 2 * \{1 - \Phi(|Y_i|)\}$, yielding $R = 932$ discoveries ($p_{(932)} = 0.001862 < 0.001864 = 0.2 \cdot 932/10^5$) with $|Y_i| > 3.111$: θ_i is declared positive for $Y_i > 3.111$ and negative for $Y_i < -3.111$. As all $\theta_i \neq 0$ this ensures directional-FDR less than 0.1. The number of simulated positive selected θ_i with negative Y_i and negative selected θ_i with positive Y_i is 56, thus the realized directional-FDR is 0.060.

The 932 selected components are displayed in Figure 1. The abscissa of the plot corresponds to Y_i, the ordinates are θ_i. The red lines are two-sided Normal 0.95 CIs: $Y_i \pm Z_{1-0.05/2}$. The Normal 0.95 CIs cover 95,089 of the 100,000 simulated θ_i, but only 610 of the 932 selected θ_i, thus the observed FCR is 0.346. The green lines are 0.05 FCR-adjusted CIs: $Y_i \pm Z_{1-0.05 \cdot 932/(2 \cdot 10^5)}$. The observed FCR for the FCR adjusted CIs is 0.046.

1.2 Selective inference in Genomic association studies

The need to correct inference for selection is widely recognized in Genome-wide association studies (GWAS). GWAS typically test association between a disease and hundreds of thousands of markers located throughout the human genome, often expressed as an odds ratio of manifesting the disease in carriers of a risk allele. Only multiplicity-adjusted significant findings are reported. This limits the occurrence of false positives, however it introduces bias into the odds ratio estimates. Analyzing 301 published studies covering 25 different reported associations, Lohmueller et al. (2003) found that for 24 associations the odds ratio in the first positive report exceeded the genetic effect estimated by meta-analysis of the remaining studies. Zollner and Pritchard (2007) suggest correcting for the selection bias by providing point estimates and CIs based on the likelihood conditional on having observed a significant association. Zhong

and Prentice (2008) further assume that in the absence of selection the log odds ratio estimator is Normally distributed. Similarly to our Bayesian analysis of the simulated example, they base their inference on a truncated normal conditional likelihood.

1.3 Parameter selection in Bayesian analysis

Berry and Hochberg (1999) comment that the Bayesian treatment of the multiplicity problem also includes decision analysis, rather than just finding posterior distributions.

Scott and Berger (2006) discuss Bayesian analysis of microarray data. The prior model for θ_i, the expectation of the log-fold change in expression of Gene i, is that $\theta_i = 0$ with probability p and $\theta_i \sim N(0, V)$ with probability $1-p$. The decision analysis performed in Scott and Berger (2006) is the discovery of the subset of active genes. Scott and Berger (2006) declare a gene active ($\theta_i \neq 0$) if the posterior expected loss of this action is smaller than the posterior expected loss of declaring the gene inactive ($\theta_i = 0$). Where the loss function for deciding that $\theta_i = 0$ is proportional to $|\theta_i|$, while the loss for erroneously deciding that $\theta_i \neq 0$ is the fixed cost of doing a targeted experiment to verify that the gene is in fact active.

In Bayesian FDR analysis of microarray data the decision analysis is also deciding which genes are active. However instead of specifying Bayes rules for selecting active genes that minimize the loss incurred by selecting inactive genes and failing to select active genes. In Efron et al. (2001), θ_i is selected if the posterior probability given y_i that $\theta_i = 0$ is less than a nominal value q. While Storey (2002, 2003) suggests using selection rules that ensure that the probability that θ_i is falsely selected is less than q.

1.4 Selection bias in Bayesian analysis

Selection is considered to have no effect on Bayesian inference. Dawid (1994) explains "Since Bayesian posterior distributions are already fully conditioned on the data, the posterior distribution of any quantity is the same, whether it was chosen in advance or selected in the light of the data." Senn (2008) reviews the disagreement between Bayesian and frequentist approaches regarding selection. He considers the example of providing inference for μ_{i^*}, the effect of the pharmaceutical associated with the largest sample mean y_{i^*}, among a class of m compounds with $Y_i \sim N(\mu_i, 4)$. He first shows that if μ_i are iid

$N(0,1)$ the posterior distribution of μ_{i^*} is $N(y_{i^*}/5, 4/5)$. He then assumes a hierarchical model in which the treatments form a compound class. The class effect is $\lambda \sim N(0, 1-\gamma^2)$ and μ_i are iid $N(\lambda, \gamma^2)$. In this case he shows that the posterior distribution of μ_{i^*} depends on the number of other compounds and their overall mean, however it is unaffected by the fact that μ_{i^*} was selected because it corresponds to the largest sample mean.

Mandel and Rinott (2009) discuss the example of a regulator trying to assess drug toxicity in a Phase I study, in which the Pharmaceutical company performs toxicity experiments on multiple drugs and drug doses but only discloses the results of successful experiments with few adverse events. $X_i \sim Binom(n, p_i)$ is the number of adverse events in each experiment. Thus the regulator bases his inference on $\{X_{T_1}, X_{T_2} \cdots\}$, where $T_j(\boldsymbol{x}) = t_j$ is the index of the j-th successful experiment. They show that if $p_1, p_2, \ldots$ are independent then the Bayesian inference of the regulator, for a "safe" drug p_{t_j}, is the same as that of the company, and it is unaffected by selection. Whereas if $p_1, p_2, \ldots$ are dependent, in particular if the company repeatedly tests the same drug until it is found safe, the Bayesian inference obtained by the regulator is affected by selection. It is different then the Bayesian inference obtained by the company, and different than the Bayesian inference he would obtain under the independence model.

1.5 Fixed and random effects in Bayesian analysis

In the Bayesian framework there can be no fixed effects since the parameters are regarded as having probability distributions. However, discussing one-way classification Box and Tiao (1973, Section 7.2) use the sampling theory terminology of fixed and random effects to distinguish between situations in which the individual means can be regarded as distinct values expected to bear no strong relationship one to another that can take take values anywhere within a wide range, and situations in which the individual means can be regarded as drawings from a distribution. Box and Tiao illustrate this distinction with the example of one-way classification of several groups of laboratory yields. In the first case the groups correspond to different methods of making a particular chemical product, while in the second case the groups correspond different batches made by the same method. The distinction only carries through to the prior model elicited for the group means. In the first case the group means are elicited flat non-informative priors. They call this model the fixed effect model. In the second case the group means are iid $N(\lambda, \sigma^2)$. This model is called the

random effect model.

1.6 Preliminary definitions and outline of the paper

Let θ denote the parameter and Y denote the data, Ω is the sample space of Y; $\pi(\theta)$ is the prior distribution of θ and $f(y|\theta)$ is the likelihood function. We define selective inference as inference provided for a function of the parameter, $h(\theta)$, that is given only if $y \in S_\Omega$ is observed, for a given subset $S_\Omega \subseteq \Omega$. For example, in our analysis of microarray data in Section 6 Y is the entire set of observed gene expression levels; $\theta = (\sigma^2, \mu)$ consists of the variances and expectations of the log-expression levels for all the genes in the array; and inference is provided for $h(\theta) = \mu_g$, the expectation of the log-fold change in expression of Gene g, only if Gene g is declared differentially expressed, by the BH procedure or the Bayesian FDR controlling selection rules introduced in Section 4.

Control over the FCR is a frequentist mechanism for providing selective inference. Notice that in Example 1.1 a random selected θ_i is covered by its FCR-adjusted CI with probability ≥ 0.95. But this frequentist selective inference mechanism suffers from several intrinsic limitations: it is impossible to incorporate prior information on the parameters; it does not provide selection adjusted point estimates or selection-adjusted inference for functions of the parameters; the selection adjustment is the same regardless of the selection criterion applied and the value of the estimator. Figure 1 suggests that the selection adjustment needed is shrinking the CIs toward 0, rather then just widening the CIs, and that smaller selection adjustments are needed for θ_i with large $|Y_i|$.

In selective inference the entire data set $Y = y$ is observed. However, as inference is provided for $h(\theta)$ only if $y \in S_\Omega$, then $Y = y$ used for providing selective inference for $h(\theta)$ is actually a realization of the joint distribution of (θ, Y), truncated by the event that $y \in S_\Omega$. Thus in order to provide Bayesian selective inference for $h(\theta)$ we define a framework for providing Bayesian inference based on the truncated distribution of (θ, Y). We call this inference selection-adjusted Bayesian (saBayes) inference (describing Bayesian selective inference a truncation problem was suggested by Bradley Efron in private communication; for a discussion on truncation see Mandel (2007) and Gelman et al. (2004) Section 7.8).

In Section 2, in order to define the components of saBayes inference: the selection-adjusted prior distribution $\pi_S(\theta)$, the selection-adjusted likelihood function $f_S(Y|\theta)$ and the selection-adjusted posterior distribution $\pi_s(\theta|y)$. We study

the effect of truncation on the marginal distribution of θ and the conditional distributions of $Y|\theta$ and $\theta|Y = y$, in a generative model, in which θ is sampled from $\pi(\theta)$ and $Y|\theta$ is sampled from $f(y|\ \theta)$. We specifically consider (θ, Y) generated by models that correspond to Box and Tiao's random effect model and fixed effect models. We also consider the case that $\pi(\theta)$ is a non informative prior, for which the generative model for θ does not apply.

In Section 3 we formally define saBayes inference, supporting our observation that saBayes inference should be used for providing Bayesian selective inference, by showing that the actions that minimize the selection-adjusted posterior expected loss are Bayes rules in selective inference. We also define a Bayesian FCR for the random effect model and explain the relation between saBayes inference and providing FCR control. In Section 4 we define the Bayesian FDR as a special case of the Bayesian FCR; present methodology for specifying selection rules that control the FDR in the random effect model; explain how this methodology can be applied to control the FDR in eBayes analysis. In Section 5 we show that the Bayesian FDR methods presented in Section 4 are generalizations of the existing Bayesian FDR methods and describe how to provide Bayes inference for selected parameters in the two group mixture model.

In Section 6 we analyze microarray data for which the level 0.10 BH procedure applied to t statistic p-values fails to discover any differentially expressed genes. While applying the level 0.10 BH procedure to p-values corresponding to hybrid frequentist/eBayes moderated t-statistics does manage to discover 245 differentially expressed genes, however it is not clear how to provide frequentist selective inference for these discoveries. We show that our level 0.05 Bayesian FDR selection rule based on the moderated t-statistic yields 1124 discoveries and that our level 0.05 Bayesian FDR selection rule based on the optimal statistic yields 1271 discoveries, and we provide Bayesian selective inference for the expected log2-fold change in expression of a specific differentially expressed gene.

2 Modelling saBayes inference

2.1 Fixed and random effects in Bayesian selective inference

The most important step in providing Bayesian selective inference is determining the way that selection acts on the parameter. A parameter is, intrinsically, either a "fixed" effect if it is generated before the data is generated and selection is

applied, a "random" effect if it is generated with the data and selection is applied to it, or a "mixed" effect if it is constructed of "fixed" and "random" effects. For example, in the microarray data analysis in Section 6 the parameters for Gene $g = 1 \cdots G$ are μ_g the expected change in expression due to the Swirl mutation and σ_g^2 the measurement error variance. Both $\mu = \{\mu_1 \cdots \mu_G\}$ and $\sigma^2 = \{\sigma_1^2 \cdots \sigma_G^2\}$ are regarded as having probability distributions. Since the values of the components of σ^2 are expected to vary according to the specific conditions of the experiment, σ^2 is a "random" effect, while μ the vector of (unknown) biological constants is a "fixed" effect, and $\theta_g = (\mu_g, \sigma_g^2)$ is "mixed" effect.

To define the components of saBayes inference, we derive the truncated distribution of θ and y in a generative model in which $\theta \sim \pi(\theta)$ and $Y|\theta \sim f(y|\theta)$, when θ is either a "fixed", "random" or "mixed" effect.

The "fixed" effect truncated sampling model. When θ is a "fixed" effect, then first θ is sampled from $\pi(\theta)$ and then selection, given by the event $S = S_\Omega \subseteq \Omega$, is applied to Y. The truncated conditional distribution of Y given θ is

$$f_S(y|\theta) \;=\; I_{S_\Omega}(y) \cdot f(y|\,\theta) / \Pr(S_\Omega|\,\theta). \tag{1}$$

But as selection is applied after θ is generated it has no affect on the truncated marginal distribution of of θ

$$\pi_S(\theta) \;=\; \pi(\theta), \tag{2}$$

thus the joint truncated distribution of (θ, Y) is given by

$$\pi_S(\theta) \cdot f_S(y|\theta). \tag{3}$$

The "random" effect truncated sampling model. In this case selection given by the event $S = \{(\theta, y) : y \in S_\Omega\}$ is applied to (θ, Y). The joint truncated distribution of (θ, Y) is the conditional density of (θ, Y) given S

$$\frac{I_{S_\Omega}(y) \cdot \pi(\theta) \cdot f(y|\,\theta)}{\int_S \pi(\theta) \cdot f(y|\,\theta) d\theta dy} \;=\; \frac{\pi(\theta) \cdot f(y|\,\theta)}{\Pr(S)}. \tag{4}$$

Notice that in this case the joint truncated density of (θ, Y) is proportional to the joint density of (θ, Y), $\pi(\theta) \cdot f(y|\theta)$. Integrating (4) over Y yields the marginal truncated distribution of θ

$$\pi_S(\theta) \;=\; \int_{S_\Omega} \frac{\pi(\theta) \cdot f(y|\,\theta)}{\Pr(S)} dy \;=\; \frac{\pi(\theta) \cdot \Pr(S_\Omega|\,\theta)}{\Pr(S)}. \tag{5}$$

Dividing (4) by (5) reveals that also in this case the truncated conditional distribution of Y given θ is $f_S(y|\ \theta)$ in (1) and that the joint truncated density of (θ, Y) can be expressed as $\pi_S(\theta) \cdot f_S(y|\ \theta)$.

The "mixed" effect truncated sampling model. We consider the "mixed" θ truncated distribution of (θ, Y) in a hierarchical generative model in which $\lambda \sim \pi_2(\lambda)$ is a "fixed" hyperparameter and $\theta|\lambda$ is a "random" effect sampled rom $\pi_1(\theta|\ \lambda)$. In this case λ is sampled and then selection, given by $S = \{(\theta, y) :\ y \in S_\Omega\}$, is applied to (θ, Y). Thus the joint truncated density of (λ, θ, Y) is

$$\frac{I_{S_\Omega}(y) \cdot \pi_2(\lambda) \cdot \pi_1(\theta|\ \lambda) \cdot f(y|\ \theta)}{\int_S \pi_1(\theta|\ \lambda) \cdot f(y|\ \theta) d\theta dy} \ = \ \frac{I_{S_\Omega}(y) \cdot \pi_2(\lambda) \cdot \pi_1(\theta|\ \lambda) \cdot f(y|\ \theta)}{\Pr(S|\ \lambda)}. \tag{6}$$

Integrating out λ in (6) yields the joint truncated density of (θ, y)

$$I_{S_\Omega}(y) \cdot f(y|\ \theta) \cdot \int \frac{\pi_2(\lambda) \cdot \pi_1(\theta|\ \lambda)}{\Pr(S|\ \lambda)} d\lambda, \tag{7}$$

and integrating out y over S_Ω yields the marginal truncated distribution of θ

$$\pi_S(\theta) \ = \ \Pr(S_\Omega|\ \theta) \int_\Lambda \frac{\pi_2(\lambda) \pi_1(\theta|\ \lambda)}{\Pr(S|\ \lambda)} d\lambda. \tag{8}$$

Again, dividing (7) by (8) yields $f_S(y|\ \theta)$ in (1) and the joint truncated density of (θ, Y) can be expressed $\pi_S(\theta) \cdot f_S(y|\theta)$.

2.2 Defining the components of saBayes inference

We will now assume that $\pi(\theta)$ is the prior distribution and $f(y|\theta)$ is the likelihood function, and use the relation between the truncated and untruncated distribution of (θ, Y) in the three generative models to define the components of saBayes inference. The selection-adjusted likelihood is defined $f_S(y|\ \theta)$ in (1), the conditional distribution of $Y|\theta$ in the three truncated sampling models; the selection-adjusted prior for "fixed", "random" and "mixed" θ is $\pi_S(\theta)$ for the corresponding truncated marginal distribution of θ given in (2), (5) and (8); the selection-adjusted posterior distribution is defined

$$\pi_S(\theta|\ y) \ = \ \frac{\pi_S(\theta) \cdot f_S(y|\ \theta)}{m_S(y)} \tag{9}$$

for $m_S(y) \ = \ \int \pi_S(\theta) \cdot f_S(y|\ \theta) d\theta$. Thus only for "random" θ the selection-adjusted posterior distribution is unaffected by selection and it is equal to the

posterior distribution $\pi(\theta|\ y)$.

Remark 2.1 Note that even though we defined the selection-adjusted posterior distribution, for 'fixed", "random" and "mixed" θ, according to the conditional distribution of θ given selection and $Y = y$. Dawid's argument, that selection has no effect on posterior distributions since conditioning on the selection event is made redundant by conditioning on $Y = y$, only applies for "random" θ, in which the selection event S is a subset of the sample space of (θ, Y). Whereas for "fixed" and "mixed" θ, for which selection is not applied to (θ, Y), $\pi_S(\theta|\ y)$ is different than $\pi(\theta|\ y)$.

Example 2.2 $\mu_1 \sim N(0, 1-\gamma^2)$ is a "fixed" effect, $\mu_2 \sim N(0, \gamma^2)$ is a "random" effect, and $Y \sim N(\mu_2 - \mu_1, 1)$. Thus for $0 \leq \gamma^2 \leq 1$ and $\theta = \mu_2 - \mu_1$, the marginal density of θ is $\pi(\theta) = \phi(\theta)$ and the conditional density of $Y|\theta$ is $f(y|\ \theta) = \phi(y - \theta)$. To illustrate the difference between the selection adjusted posterior distributions for "random", "fixed" and "mixed" effects we compute the selection adjusted posterior mean of θ for the selection rule $S_\Omega = \{y : y \geq 0\}$, for $\gamma^2 = 1$, 0 and 0.5.

For $\gamma^2 = 1$, $\theta = \mu_2$ is a "random" effect whose selection-adjusted posterior distribution, given by

$$\pi_S(\theta|\ y) \ \propto\ e^{-\frac{\theta^2}{2}} \cdot e^{-\frac{(\theta-y)^2}{2}} \ \propto\ e^{-\frac{(\theta-y/2)^2}{2\cdot(1/2)}},$$

is $N(y/2, 1/2)$ for any selection criteria. Thus $E(\theta|\ y = 1) = 0.5$.

For $\gamma^2 = 0$, $\theta = \mu_1$ is a "fixed" effect. The selection-adjusted posterior distribution is given by

$$\pi_S(\theta|\ y) \ \propto\ e^{-\frac{\theta^2}{2}} \cdot e^{-\frac{(\theta-y)^2}{2}} / \Pr(Y \geq 0|\ \theta).$$

As $\Pr(Y \geq 0|\ \theta)$ decreases in θ, the selection-adjustment stochastically decreases the posterior distribution distribution of θ, and thus $E(\theta|\ y = 1) = 0.10$.

$\gamma^2 = 0.5$ yields a "mixed" effect truncated sampling model: $\mu_1 \sim N(0, 1/2)$ is the "fixed" hyperparameter, $\theta|\mu_1$ is the $N(\mu_1, 1/2)$ "random" effect, and the selection-adjusted posterior distribution is given by

$$\pi_S(\theta|\ y) \ \propto\ e^{-\frac{(\theta-y)^2}{2}} \int e^{-\frac{\mu_1^2}{2\cdot(1/2)}} e^{-\frac{(\theta-\mu_1)^2}{2\cdot(1/2)}} / \Pr(Y \geq 0|\ \mu_1) d\mu_1.$$

As $Y|\mu_1$ is $N(\mu_1, 3/2)$ in this case the selection-adjustment is weaker, thus $E(\theta|\ y = 1) = 0.33$.

Example 2.3 Notice that Senn's example of providing inference for the most active compound is a selective inference problem in which the parameter is the vector of effects of the m pharmaceuticals $\mu = \{\mu_1 \cdots \mu_m\}$; $\mu_1 \cdots \mu_m$ are iid $N(\lambda, \gamma^2)$ for $\lambda \sim N(0, 1-\gamma^2)$; the data is $Y = \{Y_1 \cdots Y_m\}$ with $Y_i \sim N(\mu_i, 4)$; and inference is provided for $h(\mu) = \mu_i$ only if $S_\Omega = \{y : y_i = max_{j=1\cdots m} y_j\}$ occurs. Senn (2008) concludes that selection has no affect on the posterior distribution of $h(\mu)$ because in his analysis μ is a "random" effect. To show that Bayesian inference may be affected by selection, we compute the selection-adjusted posterior mean of $h(\mu) = \mu_2$ for $m = 2$ and $y = (0, 2)$, for "mixed" and "fixed" μ.

To define the "mixed" μ, we assume that λ is a "fixed" effect and $\mu|\lambda$ is a "random" effect. However, since in this example $\Pr(S_\Omega|\lambda) \equiv \Pr(S_\Omega) = 0.5$, then the "mixed" effect model truncated joint density defined in (7) reduces to the "random" effect joint density in (4). Thus in this case the conditional distribution of μ_2 is unaffected by selection. We use Expression (4) in Senn (2008) to compute the conditional mean of θ_2 for the case of "random" and "mixed" μ. For $\gamma^2 = 1$ it equals 0.4 and for $\gamma^2 = 0.5$ it equals 0.384.

The selection-adjusted joint density of μ for "fixed" λ and μ is given by

$$\pi_S(\mu_1, \mu_2|\ y = (0, 2)) \ \propto \ \frac{e^{-\frac{\lambda^2}{2\gamma^2}} \cdot e^{-\frac{(\mu_1-\lambda)^2}{2\cdot(1-\gamma^2)}} \cdot e^{-\frac{(\mu_2-\lambda)^2}{2\cdot(1-\gamma^2)}} \cdot e^{-\frac{(0-\mu_1)^2}{2\cdot 4}} \cdot e^{-\frac{(2-\mu_2)^2}{2\cdot 4}}}{Pr(Y_2 \geq Y_1|\mu_1, \mu_2)}.$$

In this case the selection adjustment increases the posterior distribution of μ values with $\mu_2 < \mu_1$, thereby stochastically decreasing the marginal posterior distribution of μ_2. For $\gamma^2 = 1$ the conditional mean of θ_2 is 0.164 and for $\gamma^2 = 0.5$ it is 0.257.

2.3 saBayes inference in the random effect model

Using the terminology suggested by Box and Tiao, we call the model for $\theta = (\theta_1 \cdots \theta_m)$ and $Y = \{Y_1 \cdots Y_m\}$, that θ_i are iid $\pi(\theta_i)$ and $Y_i|\theta_i$ are independent $f(y_i|\theta_i)$, a random effect model.

In the random effect model θ can be a "random" effect, a "fixed" effect, and even a "mixed" effect when there are iid "fixed" λ_i for which $\theta_i|\lambda_i$ are independent "random" effect. In any case the joint distribution of (θ, Y) is

$$\pi(\theta) \cdot f(y|\theta) = \Pi_{i=1}^m \pi(\theta_i) \cdot \Pi_{i=1}^m f(y_i|\theta_i). \tag{10}$$

In selective inference for $h(\theta) = \theta_i$ with $S_\Omega = \{y : y_i \in S_{marg}\}$, incorporating (10) into (3) and integrating over $\theta^{(i)} = \{\theta_1, \ldots \theta_{i-1}, \theta_{i+1}, \ldots \theta_m\}$, yields the selection adjusted joint distribution of (θ_i, Y_i) for "fixed" θ

$$\int \frac{I_{S_\Omega}(y) \cdot \pi(\theta) \cdot f(y|\theta)}{\Pr(S_\Omega|\theta)} d\theta^{(i)} = \frac{I_{S_{marg}}(y_j) \cdot \pi(\theta_j) \cdot f(y_j|\theta_j)}{\Pr(S_{marg}|\theta_i)}. \tag{11}$$

While incorporating (10) into (4) and integrating over $\theta^{(i)}$, reveals that the selection adjusted joint distribution of (θ_i, Y_i) for "random" θ is

$$\frac{I_{S_{marg}}(y_i) \cdot \pi(\theta_i) \cdot f(y_i|\theta_i)}{\Pr(S_{marg})}, \tag{12}$$

and incorporating (10) into (7) and integrating over $\theta^{(i)}$, the selection adjusted joint distribution of (θ_i, Y_i) for "mixed" θ is

$$I_{S^j_{marg}}(y_i) \cdot f(y_i|\ \theta_i) \cdot \int \frac{\pi_2(\lambda_i) \cdot \pi_1(\theta_i|\ \lambda_j)}{\Pr(S_{marg}|\lambda_i)} d\lambda_i. \tag{13}$$

2.3.1 The non-exchangeable random effect model

The non-exchangeable random effect model is a generalization of the random effect model for situations in which θ_i are distinct values expected to bear no strong relationship one to another, i.e. situations for which the non-informative prior suggested by Box and Tiao is the fixed effect model. In the non-exchangeable random effect model θ_i are independent $\pi^i(\theta_i)$ and $Y_i|\theta_i$ are independent $f(y_i|\theta_i)$. Thus the joint distribution of (θ, Y) is

$$\pi(\theta) \cdot f(y|\theta) = \Pi_{i=1}^m \pi^i(\theta_i) \cdot \Pi_{i=1}^m f(y_i|\theta_i). \tag{14}$$

The marginal distribution of (θ_i, Y_i) is

$$\pi^i(\theta_i) \cdot f(y_i|\theta_i).$$

But in selective inference for $h(\theta) = \theta_i$ with $S_\Omega = \{y : y_i \in S_{marg}\}$, the selection adjusted joint distribution of (θ_i, Y_i) for "fixed" θ is

$$\frac{I_{S_{marg}}(y_i) \cdot \pi^i(\theta_i) \cdot f(y_i|\theta_i)}{\Pr(S_{marg}|\theta_i)}. \tag{15}$$

Example 2.4 Notice that (θ, Y) in Example 1.1 were generated by the random effect model that $\theta_1 \cdots \theta_{100,000}$ are independently drawn from

$$\pi(\theta_i) \ = \ 0.9 \cdot \pi_1(\theta|\lambda = 10) + 0.1 \cdot \pi_1(\theta|\lambda = 1) \tag{16}$$

and $Y_i|\theta_i$ are independently drawn from $f(y_i|\theta_i) = \phi(y_i - \theta_i)$. Figure 1 is a scatter plot of 932 (θ_i, y_i) with $|y_i| > 3.111$; Figure 4 displays the 470 components with $y_i > 3.111$. For comparison in the comparable non-exchangeable random effect model: for $i = 1 \cdots 90000$, $\theta_i \sim \pi_1(\theta|\lambda = 10)$ and for $i = 90001 \cdots 100000$, $\theta_i \sim \pi_1(\theta|\lambda = 1)$.

It is important to note that defining θ as either a "random", "fixed" or "mixed" effect changes the truncated distribution of (θ, Y), however it has no effect on the distribution of (θ, Y) sampled Example 1.1. To observe the difference between the truncated distributions we sampled 1000 realizations of (θ, Y) from each truncated distribution for $h(\theta) = \theta_1$ with $S_\Omega = \{y : |y_1| > 3.111\}$. Figure 2 displays scatter plots of (θ_1, Y_1) from the realizations of (θ, y) with $y_1 > 3.111$. The left panel is the scatter plot for the "random" θ model. In this case the joint density of (θ_1, Y_1), given in (12), is identical to the joint density of (θ_i, Y_i) displayed in Figures 1 and 4. The right panel is the scatter plot for the "fixed" θ model with joint density given in (11). In this model $\pi_S(\theta_1|y_1)$, the selection-adjusted marginal posterior distribution of θ_1, is shrunk towards 0. For the "mixed" θ model, λ_i are iid "fixed" effects sampled from $\{10, 1\}$ with probabilities 0.90 and 0.10 and $\theta_i|\lambda_i$ are independent "random" effects with conditional density $\pi_1(\theta_i|\ \lambda_i)$. Thus the joint density of (θ_1, Y_1) given in (13) is

$$\phi(y_1 - \theta_1) \cdot \{\frac{0.9 \cdot \pi_1(\theta_j|\ \lambda_1 = 10)}{\Pr(|Y_1| > 3.111 \mid \lambda_j = 10)} + \frac{0.1 \cdot \pi_1(\theta_j|\ \lambda_1 = 1)}{\Pr(|Y_1| > 3.111 \mid \lambda_j = 1)}\}.$$

In this model the shrinking of $\pi_S(\theta_1|y_1)$ towards 0 is weaker than in the "fixed" θ model.

2.4 saBayes inference for non-informative priors

Our generative model results, regarding the effect of selection on the marginal distribution of θ, do not apply when $\pi(\theta)$ is a non-informative prior. Non-informative prior distributions are used to allow conditional analysis on θ when no prior information on θ is available (Berger 1985, Section 3.3.1). As Y also provides all the information on θ in the truncated data problem, we argue that $\pi_s(\theta)$ the prior distribution used for saBayes inference should also be a non-informative prior. We further argue that the lack of prior knowledge on θ may affect our decision to provide selective inference, but the opposite is not true – the decision to provide inference only for certain values of Y should have no effect on the non-informative prior elicited for θ. We therefore propose setting

$\pi_s(\theta) \equiv \pi(\theta)$, thus the selection-adjusted posterior distribution is given by

$$\pi_S(\theta|\ y)\ \propto\ \pi(\theta) \cdot f_S(y|\ \theta). \tag{17}$$

Which means that if θ is elicited a non-informative prior then it is treated as a "fixed" effect.

3 Selection-adjusted Bayesian inference

To formally define saBayes inference, we assume that the inference involves an action $\delta(Y)$ associated with a loss function $L(h(\theta), \delta)$. As selective inference is provided only for selected (θ, Y), $r_S(\delta)$ the expected loss incurred by $\delta(Y)$ in selective inference, which we call the saBayes risk, can be expressed as the Bayes risk for the truncated distribution of (θ, Y)

$$\begin{aligned} r_S(\delta) &= \int_{\theta\in\Theta}\int_{y\in S_\Omega} L(h(\theta), \delta(y)) \cdot \pi_S(\theta) \cdot f_S(y|\theta) dy d\theta \\ &= \int_{y\in S_\Omega} [\int_{\theta\in\Theta} L(h(\theta), \delta(y)) \cdot \pi_S(\theta|\ y) d\theta] \cdot m_S(y) dy. \end{aligned} \tag{18}$$

Thus the Bayes rules in selective inference are the actions minimizing the selection-adjusted posterior expected loss

$$\rho_S(\delta, y)\ =\ \int L(h(\theta), \delta(y)) \cdot \pi_S(\theta|\ y) d\theta,$$

and in general Bayesian selective inference should be based on $\pi_S(h(\theta)|\ y)$, the selection-adjusted posterior distribution of $h(\theta)$. Thus selection-adjusted $1-\alpha$ credible intervals for $h(\theta)$ are subsets A for which $\Pr_{\pi_S(h(\theta)|\ y)}(h(\theta) \in A) = 1-\alpha$, and the posterior mean or mode of $\pi_S(h(\theta)|\ y)$ can serve as selection-adjusted point estimators for $h(\theta)$.

Example 3.1 We provide saBayes inference for the data simulated in Example 1.1, for $h(\theta) = \theta_{12647}$ with $S_\Omega = \{y : |y_{12647}| > 3.111\}$, and for $h(\theta) = \theta_{90543}$ with $S_\Omega = \{y : |y_{90543}| > 3.111\}$. We use two prior models for θ in our analysis. In the first model θ is a "random" effect generated by the random effect model, with $\pi(\theta_i)$ in (16). In this model the saBayes posterior distribution of θ_i is proportional to the distribution of (θ_i, Y_i) in (12)

$$\pi_S(\theta_i|\ y_i)\ \propto\ \pi(\theta_i) \cdot \phi(y_j - \theta_i). \tag{19}$$

In the second model θ is generated by the non-exchangeable random effect model with unknown $\pi^i(\theta_i)$. Thus following Box and Tiao we use the flat non-informative prior $\pi^i(\theta_i) = 1$ in our analysis. The flat prior unadjusted posterior distribution of θ_i is

$$\pi(\theta_i \mid y_i) \ \propto \ \phi(y_i - \theta_i). \tag{20}$$

Whereas the non-informative prior saBayes posterior distribution of θ_i is proportional to the distribution of (θ_i, Y_i) for "fixed" θ in (11)

$$\pi_S(\theta_j \mid y_j) \ \propto \ \phi(y_j - \theta_j)/Pr(S_\Omega \mid \theta_j), \tag{21}$$

with $Pr(S_\Omega \mid \theta_j) = \Phi(-3.111 - \theta_j) + 1 - \Phi(3.111 - \theta_j)$.

Figure 3 displays the posterior distributions of θ_{12647} (left panel) and θ_{90543} (right panel). The flat prior unadjusted posterior mean and mode of θ_{12647} equal $Y_{12647} = 3.40$, the 0.95 credible interval is $[1.44, 5.36]$. The selection adjustment shrinks the posterior distributions of θ_{12647} towards 0. The "random" θ saBayes posterior distribution of θ_{12647} is bimodal with a spike at 0 and a mode at 2.40, the posterior mean is 1.68, the 0.95 credible interval is $[-0.11, 4.20]$. The flat prior saBayes posterior mode of θ_{12647} is 0.74, the posterior mean is 1.88, and the 0.95 credible interval is $[-0.04, 4.64]$.

The flat prior unadjusted posterior mean and mode of θ_{90543} equal $Y_{90543} = 5.59$, the 0.95 credible interval is $[3.63, 7.55]$. The much larger Y_{90543} produces a non-negligible likelihood only for θ_i values that correspond to almost certain selection. Thus in this case the selection adjustment is small: the flat prior saBayes posterior mode is 5.57, the posterior mean is 5.48, and the 0.95 credible interval is $[3.26, 7.52]$. The shrinking towards 0 in the "random" θ model posterior is stronger: the posterior mean and mode is 4.59 the 0.95 credible interval is $[2.62, 6.55]$.

Remark 3.2 It is important to note that as extremely unlikely values of θ with an extremely small selection probability can have a large selection-adjusted likelihood, the selection adjustment posterior distribution can be be very different than the unadjusted posterior distribution. The selection-adjusted likelihood can even be non-informative and improper – if the selection rule only includes the observed value $Y = y$ then the selection-adjusted likelihood is constant for all parameter values. Example 3.3 illustrates this phenomenon, shows how it is affected by the choice of the selection rule and that it is not unique to Bayesian

selective inference. In this paper we employ selection rules whose selection probability is minimized at $\theta = 0$ and approaches 1 for large $|\theta|$, thus the selection adjustments shrink the likelihood towards 0.

Example 3.3 To illustrate the potential non-robustness of the selection adjustment we derive the non-informative prior saBayes posterior distribution of θ_{12647}, given in (21), for an alternative one-sided selection rule $S_\Omega = \{y : y_{12647} > 3.111\}$. In this case the selection-adjusted posterior is stochastically smaller and much more diffused. The selection-adjusted posterior mode is 0.19 and the selection-adjusted posterior mean is -2.87; the 0.95 selection-adjusted credible interval is $[-15.41, 3.91]$; and an unlikely value $\theta_{12647} = -5.87$, with unadjusted likelihood $\phi(-5.87 - 3.40) = 8.73 \times 10^{-20}$ and selection probability $\Phi(-5.87 - 3.111) = 1.34 \times 10^{-19}$, has the same selection-adjusted posterior density as the unadjusted posterior mode $\theta_{12647} = 3.40$, i.e. $\pi_S(\theta_{12647} = -5.87|\ Y_{12647} = 3.40) = \pi_S(\theta_{12647} = 3.40|\ Y_{12647} = 3.40)$.

This non-robustness is not unique to Bayesian selective inference. To construct selection-adjusted frequentist 0.95 confidence intervals for θ_{12647} we begin by testing, at level 0.05 and for each value of θ_0, the null hypothesis that $\theta_{12647} = \theta_0$. The sampling distribution of $Y_{12647}|\theta_{12647} = \theta_0$ is $f_S(y_{12647}|\theta_0)$ in (1) for $\theta_{12647} = \theta_0$. Thus we reject the null hypothesis that $\theta_{12647} = \theta_0$ if y_{12647} is smaller than the 0.025 quantile or larger than the 0.975 quantile of $f_S(y_{12647}|\theta_0)$, and the 0.95 confidence interval for θ_{12647} is the set of θ_0 values for which the null hypothesis that $\theta_{12647} = \theta_0$ is not rejected for $y_{12647} = 3.40$. For the selection rule $S_\Omega = \{y : |y_{12647}| > 3.111\}$ the 0.95 confidence interval for θ_{12647} is $[-0.37, 5.03]$. While for $S_\Omega = \{y : y_{12647} > 3.111\}$ the 0.95 confidence interval for θ_{12647} is $[-9.44, 5.03]$.

3.1 FCR control in the random effect model

We define the FCR for (θ, Y) generated by the random effect model. The initial set of parameters is $\theta_1 \cdots \theta_m$. The subset of selected parameters is $\{\theta_i : y_i \in S_{marg}\}$, and a marginal confidence interval $A_{marg}(y_i)$ is constructed for each selected θ_i. For $i = 1 \cdots m$, let $R_i = I(Y_i \in S_{marg})$ and $V_i = I(Y_i \in S_{marg}, \theta_i \notin A_{marg}(Y_i))$. The indicators R_i and V_i are defined for the joint (untruncated) distribution of (θ, Y). Thus regardless of whether θ is "random" or "fixed" the

conditional density of (θ_i, Y_i) given $R_i = 1$ is

$$\frac{I_{S_{marg}}(y_i) \cdot \pi(\theta_i) \cdot f(y_i|\theta_i)}{\Pr(S_{marg})}. \tag{22}$$

In a single realization of (θ, Y), $R = \sum R_i$ is the number of selected parameters, $V = \sum V_i$ is the number of non-covering confidence intervals, and $FCP = V/max(1, R)$ is the false coverage-statement proportion. In Benjamini and Yekutieli (2005) FCR refers to a frequentist FCR, that corresponds to $E_{Y|\,\theta} FCP$ for (θ, Y) generated by a random effect model. In this paper FCR will refer to a Bayesian FCR, defined $E_{\theta,Y} FCP$. We also consider the positive FCR, $pFCR = E_{\theta,Y}(FCP|\ R > 0)$.

To explain the relation between the FCR incurred in parameter selection in the random effect model and saBayes inference, we consider $(\tilde{\theta}, \tilde{Y})$ with the same distribution as (θ, Y), but with $\tilde{\theta}$ being a "random" effect. In parameter selection the identity of the selected genes is determined according to $y_1 \cdots y_m$. Therefore constructing a marginal confidence interval for θ_i when it is selected can be expressed as providing selective inference for $h(\theta) = \theta_i$, with $S_\Omega = \{y : y_i \in S_{marg}\}$ and with θ viewed as a "random" effect. Which explains why (22) is equal to the "random" effect selection-adjusted distribution of (θ_i, Y_i), given in (12), and also implies that the conditional density of θ_i given $R_i = 1$ and $Y_i = y_i$ is equal to the "random" θ selection-adjusted posterior

$$\pi_S(\theta_i|y_i) \ \propto \ \pi(\theta_i) \cdot f(y_i|\theta_i). \tag{23}$$

As $\tilde{\theta}$ is per construction a "random" effect, the conditional probability given $R_i = 1$ and $Y_i = y_i$ that $\theta_i \notin A_{marg}(y_i)$ can be expressed as the selection-adjusted posterior expected loss in selective inference for $h(\tilde{\theta}) = \tilde{\theta}_i$ with $\tilde{S}_\Omega = \{\tilde{y} : \tilde{y}_i \in S^\Omega_{marg}\}$ for the loss function $L(\tilde{\theta}_i, A_i(\tilde{y})) = I(\tilde{\theta}_i \notin A_{marg}(\tilde{y}_i))$

$$\tilde{\rho}(\tilde{y}) \ = \ \int I(\tilde{\theta}_i \notin A_{marg}(\tilde{y}_i)) \cdot \pi(\tilde{\theta}_i|\ \tilde{y}_i) d\tilde{\theta}_i,$$

and the conditional probability given that $R_i = 1$ that $\theta_i \notin A_{marg}(y_i)$ is the corresponding saBayes risk

$$\tilde{r}_S \ = \ E_{m_S(\tilde{y}_i)} \tilde{\rho}(\tilde{y}_i). \tag{24}$$

Proposition 3.4 *The pFCR in the random effect model is equal to the "random" θ saBayes risk $\tilde{r}_S$. In particular, if $A_{marg}(\tilde{y}_i)$ are $1-\alpha$ credible intervals for θ_i based on $\pi_S(\theta_i|y_i)$ in (23) then $pFCR = \alpha$.*

Proof. In the random effect model R_i are independent and $\{V_i : R_i = 1\}$ are mutually independent with $\Pr(V_i = 1|R_i = 1) = \tilde{r}_S$. Thus for each value of $R = k$, $V \sim Binom(k, \tilde{r}_S)$, and conditioning on $R > 0$ yields $pFCR = \tilde{r}_S$. Lastly, for $1-\alpha$ selection-adjusted credible intervals based on $\pi_S(\theta_i|y_i)$, $\tilde{r}_S = \tilde{\rho}(\tilde{y}_i) \equiv \alpha$. ¶

Remark 3.5 We have shown that in the random effect model, regardless of whether θ is "random", "fixed" or "mixed", the pFCR equals the "random" θ saBayes risk. As pFCR $\geq$ Bayesian-FCR the "random" θ saBayes risk can serve as a conservative estimate for Bayesian-FCR. In particular, for large R the sampling dispersion of FCP and of V/ER is small, thus the FCP, Bayesian-FCR, frequentist-FCR, pFCR and also EV/ER, we discuss in the context specifying selection rules in the non-exchangeable random effect model, are almost the same.

Remark 3.6 Recall that if $\pi(\theta_i)$ is a noninformative prior then the selection adjusted posterior distribution for "random" θ is actually the "fixed" θ selection adjusted posterior

$$\pi_S(\theta_i|y_i) \propto \pi(\theta_i) \cdot f(y_i|\theta_i)/\Pr(S_{marg}|\theta_i). \tag{25}$$

As credible intervals based on non-informative priors are expected to provide approximate coverage probability, when $\pi(\theta_i)$ is a non-informative prior then $1-\alpha$ credible intervals based on $\pi_S(\theta_i|y_i)$ in (25) yield $\tilde{\rho}(y_i) \approx \alpha$. Thus Proposition 3.4 implies that for non informative priors the "fixed" θ marginal $1-\alpha$ credible intervals yield approximate level α FCR control.

Example 3.7 Figure 4 displays (θ_i, y_i) generated in Example 1.1 with $y_i > 3.111$. The red and green dashed curves are the 0.95 confidence intervals from Figure 1. The red curves also correspond to the 0.95 credible intervals for θ_i for the flat prior unadjusted posterior (20). The blue curves are the 0.95 saBayes credible intervals for the flat prior selection-adjusted posterior in (21), and the light blue curves are the 0.95 saBayes credible intervals for the "random" θ selection-adjusted posterior in (19).

According to Proposition 3.4 the pFCR for "random" θ 0.95 saBayes credible intervals constructed for selected (θ_i, y_i) is 0.05. In the simulation the FCP for the 932 selected θ_i was 0.047. As the flat prior unadjusted credible intervals

are 0.95 frequentist confidence intervals, we expect the coverage proportion for all $100,000$ θ_i to be close to 0.95. In Example 1.1 we have seen that these CIs cover $95,089$ of the $100,000$ θ_i. From a Bayesian perspective these are equal tail credible intervals based on minimally-informative prior known to provide good frequentist performance (Carlin and Louis, 1996, Section 4.3). We have also seen that the FCP for the 932 selected parameters is 0.346. Benjamini and Yekutieli (2005) explain this phenomenon from a frequentist perspective. Remark 3.6 offers a Bayesian explanation: in order to provide approximate FCR control for non informative priors the credible intervals should be based on the "fixed" θ selection adjusted posterior in (20), rather than the "random" θ selection adjusted posterior in (19). And indeed, the FCP of the credible intervals based on (20) was 0.040.

4 Specifying FDR controlling selection rules in the random effect model

In this section we present methods for specifying selection rules in cases where the primary goal of the experiment is making statistical discoveries. As in Section 3.1, we assume that (θ, Y) are generated by the random effect model; θ_i is selected if $y_i \in S_{marg}$; and the inference provided for θ_i if it is selected is declaring that it is in $A_{marg}(y_i)$. However now $A_{marg}(y_i)$ is an event that corresponds to making a statistical discovery regarding θ_i. In Senn's example of providing inference for the most active compound, the statistical discovery that corresponds to selecting θ_i is declaring that $\theta_i > max_{j \neq i}\theta_j$. While in Genome-wide association studies the selected parameters are odds ratio between diseases and genetic markers that are found to be either greater than 1 or smaller than 1.

Once declaring $\theta_i \in A_{marg}(y_i)$ corresponds to making a statistical discovery, R becomes the number of discoveries, V becomes the number of false discoveries, $V/max(1,R) = FDP$ is the false discovery proportion, and $FCR = FDR$. Thus Proposition 3.4 yields the following result.

Corollary 4.1 *In the random effect model the $pFDR$ is equal to the saBayes risk for "random" θ in selective inference for $h_i(\theta) = \theta_i$ with $S_\Omega = \{y : y_i \in S_{marg}\}$ and loss function $L(\theta_i, A_{marg}) = I(\theta_i \notin A_{marg}(y_i))$.*

Thus to ensure level α FDR control, when considering $S_{marg} = \{y_i : T(y_i) \leq s\}$, we suggest choosing s for which the "random" θ saBayes risk is q.

Furthermore since for "random" θ, for which posterior distributions are unaffected by selection, the posterior expected loss is

$$\tilde{\rho}(\tilde{y}_i) \;=\; \int I(\tilde{\theta}_i \notin A_{marg}(\tilde{y}_i)) \cdot \pi(\tilde{\theta}_i |\; y_i) d\tilde{\theta}_i,$$

and the truncated marginal distribution of y_i is

$$m_s(\tilde{y}_i) = \frac{I(\tilde{y}_i \in S_{marg}) \cdot \int \pi(\tilde{\theta}_i) f(\tilde{y}_i|\tilde{\theta}_i) d\tilde{\theta}_i}{\int I(\tilde{y}_i \in S_{marg}) \cdot \int \pi(\tilde{\theta}_i) f(\tilde{y}_i|\tilde{\theta}_i) d\tilde{\theta}_i d\tilde{y}_i}.$$

For any S_{marg} the saBayes risk in (18) can be expressed

$$\begin{aligned} \tilde{r}_S(\delta) &= \frac{\int I(\tilde{y}_i \in S_{marg}) \cdot \rho(\tilde{y}_i) \cdot \int \pi(\tilde{\theta}_i) f(\tilde{y}_i|\tilde{\theta}_i) d\tilde{\theta}_i d\tilde{y}_i}{\int I(\tilde{y}_i \in S_{marg}) \cdot \int \pi(\tilde{\theta}_i) f(\tilde{y}_i|\tilde{\theta}_i) d\tilde{\theta}_i d\tilde{y}_i} \\ &= \frac{\int I(\tilde{y}_i \in S_{marg}) \cdot \tilde{\rho}(\tilde{y}_i) \cdot \tilde{m}(\tilde{y}_i) d\tilde{y}_i}{\int I(\tilde{y}_i \in S_{marg}) \cdot \tilde{m}(\tilde{y}_i) d\tilde{y}_i}, \end{aligned} \tag{26}$$

for $\tilde{m}(y_i) = \int \tilde{\pi}(\tilde{\theta}_i) f(\tilde{y}_i |\; \tilde{\theta}_i) d\tilde{\theta}_i$. Thus as the denominator in (26) is the probability that θ_i is selected, Corollary 4.1 and Expression (26) for the "random" θ the saBayes risk yield the following Neyman-Pearson Lemma type result.

Corollary 4.2 *$S_{marg} = \{y_i : \tilde{\rho}(y_i) \leq s\}$ has the largest selection probability of all selection rules with the same pFDR.*

Another option is to use $\tilde{\rho}(y_i)$ to directly specify the selection rule, by defining

$$S_{marg} = \{y_i : \tilde{\rho}(y_i) \leq q\}. \tag{27}$$

Notice that unlike the continuum of possible credible intervals that can be constructed for θ_i, the number of possible discoveries that can be made regarding θ_i is finite. In particular, when there is only a single possible discovery for all values of y_i, i.e. $A_{marg}(y_i) \equiv A_{marg}$, then expressing the "random" θ saBayes risk corresponding to this discovery

$$\begin{aligned} \tilde{r}_S &= \int \int_{\tilde{y}_i \in S_{marg}} I(\tilde{\theta}_i \notin A_{marg}) \cdot \frac{\tilde{\pi}(\tilde{\theta}_i) \cdot f(\tilde{y}_i|\tilde{\theta}_i)}{\Pr(S_{marg})} d\tilde{y}_i d\tilde{\theta}_i \\ &= \int I(\tilde{\theta}_i \notin A_{marg}) \cdot \frac{\tilde{\pi}(\tilde{\theta}_i) \Pr(S_{marg}|\tilde{\theta}_i)}{\Pr(S_{marg})} d\tilde{\theta}_i \\ &= \int I(\tilde{\theta}_i \notin A_{marg}) \cdot \pi_S(\tilde{\theta}_i) d\tilde{\theta}_i, \end{aligned} \tag{28}$$

for $\pi_S(\tilde{\theta}_i) = \tilde{\pi}(\tilde{\theta}_i) \cdot \Pr(S_{marg}|\ \tilde{\theta}_i)/\Pr(S_{marg})$ the "random" θ selection-adjusted prior density derived in (5), yields the following result.

Corollary 4.3 *If $A_{marg}(y_i) \equiv A_{marg}$ then the pFDR is equal to the "random" θ selection-adjusted prior probability that $\theta_i \notin A_{marg}$.*

4.1 Specifying FDR controlling selection rules in the non-exchangeable random effect model

In this subsection, (θ, Y) is generated by the non-exchangeable random effect model, θ_i is selected if $y_i \in S_{marg}$, and the inference provided for selected θ_i is the discovery that $\theta_i \in A_{marg}(y_i)$. Let $A^1_{marg} \cdots A^D_{marg}$ denote the D possible discoveries that can be made on θ_i. For $d = 1 \cdots D$, let R^d denote the number of discoveries of A^d_{marg} and let V^d denote the number of false discoveries of A^d_{marg}. The results in this section are derived under the assumption that $A_{marg}(y_i) \equiv A_{marg}$. However as $ER = ER^1 + \cdots + ER^D$ and $EV = EV^1 + \cdots + EV^D$, they can be easily extended for the case of $D > 1$. To derive the results in this section we consider $(\tilde{\theta}, \tilde{Y})$ as before, but with marginal prior density $\tilde{\pi}(\theta_i) = \sum_{i=1}^m \pi^i(\theta_i)/m$.

Lemma 4.4 *For any subset B, $W_i = I(y_i \in S_{marg}, \theta_i \notin B)$ and $\tilde{W}_i = I(\tilde{y}_i \in S_{marg}, \tilde{\theta}_i \notin B)$*

$$E\sum_{i=1}^m W_i = E\sum_{i=1}^m \tilde{W}_i.$$

Proof.

$$\begin{aligned}
E\sum_{i=1}^m W_i &= \sum_{i=1}^m \Pr(Y_i \in S_{marg}, \theta_i \notin B) \\
&= \sum_{i=1}^m \int_{\theta_i \notin B} \int_{y_i \in S_{marg}} \pi^i(\theta_i) \cdot f(y_i|\theta_i) dy_i d\theta_i \\
&= \sum_{i=1}^m \int_{\theta_1 \notin B} \int_{y_1 \in S_{marg}} \pi^i(\theta_1) \cdot f(y_1|\theta_1) dy_1 d\theta_1 \\
&= m \cdot \int_{\theta_1 \notin B} \int_{y_1 \in S_{marg}} \sum_{i=1}^m \pi^i(\theta_1)/m \cdot f(y_1|\theta_1) dy_1 d\theta_1 \\
&= m \cdot \int_{\theta_1 \notin B} \int_{y_1 \in S_{marg}} \tilde{\pi}(\theta_1) \cdot f(y_1|\theta_1) dy_1 d\theta_1 = E\sum_{i=1}^m \tilde{W}_i
\end{aligned}$$

¶

For $B = \emptyset$, $\sum_{i=1}^{m} W_i$ is the number of discoveries R, while for $B = A_{marg}$, $\sum_{i=1}^{m} W_i$ is the number of false discoveries. Therefore Lemma 4.4 implies that EV, ER, thus also EV/ER, for (θ, Y) and for $(\tilde{\theta}, \tilde{Y})$ are the same. According to Corollary 4.3 for $(\tilde{\theta}, \tilde{Y})$ the pFDR equals $\tilde{r}_S$. Thus as $FDR \leq pFDR$, and $FDR \approx EV/ER$ is approximately the same for (θ, Y) and for $(\tilde{\theta}, \tilde{Y})$, we get the following result.

Corollary 4.5 *In the non-exchangeable random effect model selecting θ_i if $y_i \in S_{marg}$ yields approximate level $\tilde{r}_S$ FDR control.*

To define a general method for specifying FDR controlling selection rules for (θ, Y) generated by the non-exchangeable random effect model with unknown marginal priors, notice that applying empirical Bayes methods to $y_1 \cdots y_m$ actually estimates $\tilde{\pi}(\theta_i)$, the mixture of the (unknown) marginal densities of $\theta_1 \cdots \theta_m$. Combining this with Corollary 4.5 implies that the FDR of any selection rule can be approximated by $\tilde{r}_S$ computed by treating (θ, Y) as if it was generated by the random effect model and using eBayes estimate of $\tilde{\pi}(\theta_i)$. Furthermore, as $ER = E\tilde{R}$ and $E\tilde{R} = m \cdot \Pr(\tilde{y}_i \in S_{marg})$, then also in the non-exchangeable random effect model the selection rule $S_{marg} = \{y_i : \tilde{\rho}(y_i) \leq s\}$, yields the maximal ER of all S_{marg} with the same $\tilde{r}_S$.

Definition 4.6 Algorithm for specifying level q FDR controlling selection rules in the non-exchangeable random effect model:

1. Apply eBayes to $y_1 \cdots y_m$ to produce $\tilde{\pi}(\theta_i)$.

2. Use $\tilde{\pi}(\theta_i)$ to compute $\tilde{r}_S$ for any given selection rule.

3a. To specify a level q FDR controlling selection rule of the form $S_{marg} = \{y : T(y_i) \leq s\}$, for a given statistic $T(y_i)$, find s for which $\tilde{r}_S = q$.

3b. The level q FDR controlling selection rule yielding the maximal expected number of discoveries is $S_{marg} = \{y : \tilde{\rho}(y_i) \leq s\}$ with s for which $\tilde{r}_S = q$.

Example 4.7 In Example 1.1 selection is associated with $D = 2$ directional discoveries. According to Corollary 4.1 the pFDR for the selection rule $|y_i| \geq s$ is equal to the "random" θ saBayes risk for the loss function $I(sign(\theta_i) \neq sign(y_i))$

$$E_{m_S(y)}\{\ I(y < -a) \cdot \Pr_{\pi_S(\theta|\ y)}(\theta > 0)\ +\ I(y > a) \cdot \Pr_{\pi_S(\theta|\ y)}(\theta < 0)\ \}. \qquad (29)$$

Recall that $|y_i| > 3.111$ was used to ensure that the directional-FDR is less than 0.1. For $s = 3.111$ the saBayes risk (29) is 0.070, whereas setting $s = 2.915$ yields the selection criterion for which the saBayes risk is 0.10. The posterior expected loss corresponding to the directional-FDR is

$$\tilde{\rho}(y_i) = \Pr_{\pi(\theta|\, y)}(sign(\theta_i) \neq sign(y_i)).$$

Notice that in this example $\tilde{\rho}(y_i)$ increases in $|y_i|$, thus $|y_i| \geq 2.915$ is the $\tilde{r}_S = 0.10$ selection rule yielding the maximal expected number of discoveries. For $y_i \geq 0$, $\tilde{\rho}(y_i)$ is the conditional probability given y_i that $\theta_i < 0$. $\tilde{\rho}(0) = 0.5$, $\tilde{\rho}(3.111) = 0.176$, and $\tilde{\rho}(3.472) = 0.10$. Thus $|y_i| \geq 3.472$ is the selection criterion suggested in (27) for $q = 0.10$.

To illustrate the results on the non-exchangeable random effect model, we evaluated EV, ER and the directional-FDR in $n = 10^5$ replications of the random effect model simulated in Example 1.1 and the comparable non-exchangeable random effect model described in Example 2.4. In both models the mean number of discoveries was 919.9 (s.e. < 0.07), the mean number of false discoveries was 64.4 (s.e. < 0.03), and the mean directional-FDP was 0.070 (s.e. < 0.00003).

5 The relation between saBayes inference and Bayesian FDR methods

The term Bayesian FDR methods refers to the multiple testing procedures presented in Efron et al. (2001) and Storey (2002, 2003) for the following two group mixture model. H_i, $i = 1 \cdots m$, are iid $Bernoulli(1 - \pi_0)$ random variables. $H_i = 0$ corresponds to a true null hypothesis, while $H_i = 1$ corresponds to a false null hypothesis. Given $H_i = j$, Y_i is independently drawn from f_j, for $j = 0, 1$.

The positive FDR (pFDR) corresponds to a rejection region Γ. It is defined $E(V/R|\ R > 0)$ where R is the number of $y_i \in \Gamma$, and V is the number of $y_i \in \Gamma$ with $H_i = 0$. Storey proves that

$$pFDR(\Gamma) = Pr(H_i = 0|Y_i \in \Gamma) \tag{30}$$

$$= \frac{\pi_0 \cdot Pr(Y_i \in \Gamma|H_i = 0)}{\pi_0 \cdot Pr(Y_i \in \Gamma|Y_i = 0) + (1 - \pi_0) \cdot Pr(Y_i \in \Gamma|H_i = 1)}, \tag{31}$$

with $Pr(Y_i \in \Gamma|H_i = j) = \int_{y_i \in \Gamma} f_j(y_i)dy_i$. For the multiple testing procedure each null hypothesis is associated with a rejection region Γ_i, determined by y_i;

the pFDR corresponding to Γ_i, called the q-value, is computed; and the null hypothesis $H_i = 0$ is rejected if q-value $\leq q$. The local FDR is defined in Efron et al. (2001) as the conditional probability given $Y_i = y_i$ that $H_i = 0$

$$fdr(y_i) = \frac{\pi_0 \cdot f_0(y_i)}{\pi_0 \cdot f_0(y_i) + (1 - \pi_0) \cdot f_1(y_i)}.$$

The multiple testing procedure based on the local FDR is reject $H_i = 0$ if $fdr(Y_i) \leq q$.

Notice that Bayesian FDR methods can be expressed as a special case of the FDR controlling selection rules presented in the previous section, in which the components of the parameter vector are dichotomous. The parameter is $H = (H_1 \cdots H_m)$, and (H, Y) are generated by a random effect model: the marginal distribution of H_i is $\pi(H_i = j) = (1-\pi_0)^j \cdot \pi_0^{(1-j)}$, f_j is the likelihood, H_i is selected if $y_i \in \Gamma$ and selection is associated with declaring $H_i = 1$. Notice also that Expression (31) is a special case of Expression (28): it is the "random" effect saBayes risk for the loss function $I(H_i = 0)$, expressed as the selection-adjusted prior distribution of making a a false discovery

$$\pi_\Gamma(H_i = 0) \ \propto \ \pi(H_i = 0) \cdot \Pr(Y_i \in \Gamma |\ H_i = 0).$$

Thus the equality in (30) proven by Storey is a special case of Corollary 4.3. The local FDR is the "random" θ selection-adjusted posterior expected loss, thus the multiple testing procedure based on the local FDR is a special case of the selection rule in (27). Lastly, the relation between the local FDR and the pFDR, $pFDR = E_{y \in \Gamma} fdr(y)$, follows from the definition of the saBayes risk in (18).

Bayesian FDR methods are valid regardless of whether H is a "random" or "fixed" effect. However in selective inference for $h(H) = H_i$, the selection-adjusted posterior probability that $H_i = 0$ for a "random" H is equal to the local fdr. Whereas if H is a "fixed" effect, or if π_0 is the non-informative prior probability that $H_i = 0$, then the selection-adjusted posterior distribution that $H_i = 0$ is

$$\frac{\pi_0 \cdot f_\Gamma(y_i | H_i = 0)}{\pi_0 \cdot f_\Gamma(y_i | H_i = 0) + (1 - \pi_0) \cdot f_\Gamma(y_i | H_i = 1)},$$

for $f_\Gamma(y_i |\ H_i = j) = f_j(y_i) / \Pr(y_i \in \Gamma |\ H_i = j)$ the selection-adjusted likelihood.

6 Analysis of microarray data

We analyze the Dudoit and Yang (2003) swirl data set. The data includes 4, 8448 gene arrays, comparing RNA from Zebrafish with the swirl mutation to

RNA from wild-type fish. For Gene g, $g = 1 \cdots 8448$, the parameters are μ_g the expected log2-fold change in expression due to the swirl mutation, and σ_g^2 the variance of the log2-fold change in expression.

In our analysis we assume that (θ, Y) are generated by a non-exchangeable random effect model. σ_g^2 are iid "random" effects with scaled inverse chi-square marginal prior density $\pi(\sigma_g^2)$ whose hyper-parameters, $s_0^2 = 0.052$ and $\nu_0 = 4.02$, were derived by applying the R LIMMA package (Smyth, 2005) *eBayes* function to the sample variances. μ_g are distinct independent "fixed" effects, that are elicited flat non-informative priors, $\pi_{ni}(\mu_g) \propto 1$. For assessing the FDR of the selection rules we use the eBayes prior

$$\tilde{\pi}(\mu_g) \;=\; 8.5 \cdot \exp(-8.5 \cdot |\mu_g|)/2,$$

that provided a good fit to the empirical distribution of $\bar{y}_1 \cdots \bar{y}_{8448}$. Given μ_g and σ_g, s_g^2 the sample variances are independent $\sigma_g^2 \chi_3^2/3$, and $\bar{y}_g$ the observed mean log2 expression ratios are independent $N(\mu_g,\ \sigma_g^2/4)$. Thus the marginal likelihood is given by

$$f(\bar{y}_g, s_g^2|\ \mu_g, \sigma_g^2) \;\propto\; \sigma_g^{-4} \exp\{-\frac{1}{2\sigma_g^2}[3s_g^2 + 4(\mu_g - \bar{y}_g)^2]\}. \tag{32}$$

Our goal in the analysis is to specify a selection rule for which the mean directional error in declaring selected genes with $\bar{y}_g > 0$ over-expressed and declaring selected genes with $\bar{y}_g < 0$ under-expressed is less than 0.05, and to provide inference for the change in expression of selected genes.

6.1 Specifying the selection rules

In the first part of our analysis we use the level $q = 0.10$ BH procedure to discover differentially expressed genes; assess the directional-FDR of the selection rule specified by the BH procedure; compare its performance to the level $q = 0.10$ Bayesian FDR controlling selection rule based on moderated t statistics and the most powerful level $q = 0.10$ Bayesian FDR controlling selection rule based, constructed according to the algorithm defined in 4.6.

LIMMA implements a hybrid classical/Bayes approach in which μ_g are assumed to be unknown constants while σ_g^2 are iid $\pi(\sigma_g^2)$. The moderated t statistic is defined $\tilde{t}_g = \bar{y}_g/(\tilde{s}_g/2)$, for $\tilde{s}_g^2 = (\nu_0 s_0^2 + 3s_g^2)/(\nu_0 + 3)$ the posterior mean of $\sigma_g^2|s_g^2$. As $\tilde{s}_g^2/\sigma_g^2 \sim \chi_{\nu_0+3}^2/(\nu_0+3)$, $(\bar{y}_g - \mu_g)/(\tilde{s}_g/2)$ are $(\nu_0 + 3)$ degree of freedom t random variables. Thus the p-values LIMMA provides to test the null hypotheses of non-differential expression are $\tilde{p}_g = 2 \cdot (1 - F_{\nu_0+3}(|\tilde{t}_g|))$, where F_ν

is the ν degree of freedom t cdf. Applied at level $q = 0.10$ to the 8448 p-values the BH procedure yielded 245 discoveries, corresponding to the rejection region $|\tilde{t}_g| > 4.479$. The observed mean log2 expression ratios and sample standard deviations of the 8448 genes are drawn in Figure 5. The BH discoveries are the 245 observations beneath the solid blue curve $|\tilde{t}_g| = 4.479$. To see why this rejection region corresponds to 0.05 directional FDR control notice that for all μ_g, the probability of a directional error is less than $1 - F_{\nu_0+3}(4.479)$; thus $12.08 = 8448 \cdot (1 - F_{\nu_0+3}(4.479))$ is a conservative estimate for the number of false directional discoveries, and $0.049 = 12.08/245$ is a conservative estimate for the directional FDR.

For comparison, the frequentist treatment of this problem would be testing the null hypotheses of non-differential expression by 3 degree of freedom test statistics $t_g = \bar{y}_g/(s_g/2)$. Since the 3 degree of freedom t-distribution has heavier tails, $F_3^{-1}(1-0.1/(2\cdot 8448)) = 57.10$ while $max(|t_g|)$ is only 27.90. Thus applying the level $q = 0.1$ BH to $p_1 \cdots p_{8448}$, with $p_g = 2\cdot(1-F_3(|t_g|))$, yields 0 discoveries.

In order to assess the directional FDR we derive the "random" θ saBayes posterior distribution

$$\tilde{\pi}_S(\mu_g, \sigma_g^2|\bar{y}_g, s_g) \;=\; \frac{I((\bar{y}_g, s_g^2) \in S_{marg}) \cdot \tilde{\pi}(\mu_g, \sigma_g^2) \cdot f(\bar{y}_g, s_g|\, \mu_g, \sigma_g^2)}{\Pr((\bar{y}_g, s_g^2) \in S_{marg})}, \qquad (33)$$

for the eBayes prior distribution $\tilde{\pi}(\mu_g, \sigma_g^2) = \tilde{\pi}(\mu_g) \cdot \pi(\sigma_g^2)$. We then integrate out σ_g^2 in (33) to derive $\tilde{\pi}_S(\mu_g|\bar{y}_g, s_g)$ the marginal "random" θ saBayes posterior distribution of μ_g, and the "random" θ posterior expected loss corresponding to directional errors

$$\tilde{\rho}(\bar{y}_g, s_g^2) \;=\; \int I\{\mu_g \neq sign(\bar{y}_g)\} \cdot \tilde{\pi}(\mu_g|\bar{y}_g, s_g^2\,)d\mu_g,$$

and use it to numerically compute the "random" θ saBayes risk corresponding to the directional FDR

$$\tilde{r}_S(S_{marg}) \;=\; E_{m_S(\bar{y}_g, s_g^2)}(\tilde{\rho}(\bar{y}_g, s_g^2)),$$

for

$$m_S(\bar{y}_g, s_g) \;=\; \frac{I((\bar{y}_g, s_g^2) \in S_{marg}) \cdot \tilde{\pi}(\mu_g, \sigma_g^2) \cdot f(\bar{y}_g, s_g|\, \mu_g, \sigma_g)}{\int I((\bar{y}_g, s_g^2) \in S_{marg}) \cdot \tilde{\pi}(\mu_g, \sigma_g^2) \cdot f(\bar{y}_g, s_g|\, \mu_g, \sigma_g) d\mu_g d\sigma_g}.$$

$\tilde{r}_S$ for $|\tilde{t}_g| > 4.479$ the $q = 0.10$ BH procedure (solid blue curve in Figure 5) is 0.024. While $|\tilde{t}_g| > 2.64$ (dashed blue curve in Figure 5) is the moderated t selection rule with $\tilde{r}_S = 0.05$. It yields 1124 discoveries. The green curves

in Figure 5 correspond to the selection rules $\tilde{\rho}(\bar{y}_g, s_g^2) < s$. The solid curve corresponds to the selection rule with $s = 0.05$, that yields 559 discoveries. The dashed curve corresponds to the selection rule with $s = 0.088$, for which $\tilde{r}_S = 0.05$. This is the selection rule that yields the maximal expected number of discoveries of all selection rules with $\tilde{r}_S = 0.05$. In this case it yields 1271 discoveries.

6.2 Providing saBayes inference

In the second part of our analysis we provide saBayes inference for μ_{6239}, the expected log2-fold change in expression due to the swirl mutation for Gene number 6239. Note that in the hybrid classical/Bayes approach it is not clear how to apply the Benjamini and Yekutieli (2005) frequentist FCR adjustment. The statistics for this gene (marked by the red plus sign in Figure 5) are $\bar{y}_{6239} = -0.435$ and $s^2_{6239} = 0.0173$ thus $\tilde{t}_{6239} = -4.51$.

The marginal posterior distributions of μ_{6239} are drawn in Figure 6. The black curve corresponds to the non-informative prior unadjusted posterior

$$\pi(\mu_g, \sigma_g^2 | \bar{y}_g, s_g\ ) \ \propto\ \pi_{ni}(\mu_g) \cdot \pi(\sigma_g^2) \cdot f(\bar{y}_g, s_g |\ \mu_g, \sigma_g^2),$$

for which $(\mu_{6239} - \bar{y}_{6239})/(\tilde{s}_{6239}/2) \sim t_{7.02}$. The posterior mean and mode equal $\bar{y}_{6239} = -0.435$, the 0.95 credible interval for μ_{6239} is $[-0.61, -0.21]$, the posterior probability that $\mu_{6239} > 0$ and a directional error is committed is 0.0014. The green curve corresponds to $\tilde{\pi}_S(\tilde{\mu}_{6239} | \bar{y}_{6239}, s_{6239}\ )$. Its posterior mode is -0.36, the posterior mean is -0.31, the 0.95 credible interval is $[-0.54, -0.01]$, and the posterior probability that $\mu_{6239} > 0$ is 0.020.

As μ_g is elicited a non-informative prior and σ_g^2 is a "random" effect, then the selection-adjusted posterior distribution of (μ_g, σ_g^2) is proportional to the joint truncated distribution in (6), with μ_g substituting the "fixed" λ and σ_g^2 substituting the "random" θ,

$$\pi_S(\mu_g, \sigma_g^2 | \bar{y}_g, s_g\ ) \ \propto\ \pi(\sigma_g^2) \cdot \pi_{ni}(\mu_g) \cdot f(\bar{y}_g, s_g |\ \mu_g, \sigma_g^2) / \Pr(|\tilde{t}_g| > a \mid \mu_g). \quad (34)$$

SaBayes inference for μ_{6239} is based on $\pi_S(\mu_g | \bar{y}_g, s_g)$, the marginal selection adjusted posterior of μ_{6239}, derived by integrating out σ_g from (34). The solid blue curve is $\pi_S(\mu_g | \bar{y}_g, s_g\ )$ for the selection rule $|\tilde{t}_g| > 4.479$. Its posterior mode is -0.278, the posterior mean is -0.257, the 0.95 credible interval is $[-0.54, 0.02]$, and the posterior probability that $\mu_{6239} > 0$, and thus the Gene was erroneously declared under-expressed, is 0.038. The dashed blue curve

corresponds to $|\tilde{t}_g| > 2.64$. In this case the shrinking towards 0 is weaker: the posterior mode is -0.419, the posterior mean is -0.367, the 0.95 credible interval is $[-0.63, -0.02]$, and the posterior probability that $\mu_{6239} > 0$ is 0.017.

7 Discussion

We have shown that selective inference adds an arbitrary element to Bayesian analysis. However it is important to note that the selection rule is determined before the data is observed, and once the selection rule is determined the entire process of providing saBayes inference is fully specified and is carried out the same way as Bayesian inference. The notable exception is eBayes methods in which the data is used twice in the analysis, first to elicit the prior distribution and possibly to specify the selection rule, and then to produce posterior distributions.

Our method of controlling the Bayesian FDR corresponds to the fixed rejection region approach presented in Yekutieli and Benjamini (1999), that consists of estimating the FDR in a series of nested fixed rejection regions and choosing the largest rejection region with estimated FDR less than q. However, as the pFDR of any selection rule S_{marg} can be expressed as a saBayes risk, the problem of controlling the Bayesian FDR in the random effect and non-exchangeable random effect models is reduced into a Bayesian decision problem of finding the "optimal" selection rule with $\tilde{r}_S \leq q$. Our Bayesian FDR controlling methods can, in principle, provide tight FDR control for any discovery event $A_{marg}(y_i)$. Whereas frequentist FDR controlling methods may provide tight FDR control when the discovery is rejecting a simple null hypothesis, but as illustrated by the performance of the BH procedure in controlling the directional-FDR, can only bound the FDR when the discoveries are rejecting composite null hypotheses.

In general, the price paid by using stricter selection rules is reduction in the information the data provides for selective inference. Example 3.3 suggests that when specifying selection rules, in addition to the tradeoff between allowing too many false (or wasteful) discoveries and failing to make enough discoveries, it may also be advisable to take into account the quality of the inference provided for selected parameters.

Lastly, even though we discussed selection rules that control the FDR incurred when selecting a subset of parameters and used either non-informative priors or random-effect priors. Our main result, that Bayesian inference for "fixed" and "mixed" effects must be corrected for selection, also applies when

the prior distribution is elicited according to prior knowledge and regardless of why selection is applied.

References

[1] Benjamini Y., Hochberg Y. (1995) "Controlling the False Discovery Rate: a practical and powerful approach to multiple testing" *Journal of the Royal Statistical Society, Series B*; **57** (1): 289-300.

[2] Benjamini Y., Yekutieli D. (2005) "False Discovery Rate-Adjusted Multiple Confidence Intervals for Selected Parameters" *Journal of the American Statistical Association*, **100**, 71-81.

[3] Berger J.O. (1985) *Statistical Decision Theory and Bayesian Analysis*, Springer Series in Statistics.

[4] Berry, D. A., Hochberg, Y. (1999) "Bayesian perspectives on multiple comparisons" *J. Statist. Plann. Inference*, **82**, 215227.

[5] Box G.E.P., Tiao G. C. (1992) *Bayesian inference in statistical analysis*, Wiley Classics Library Edition.

[6] Carlin B.P., Louis T.A. (1996) *Bayes and Empirical Bayes Methods for Data Analysis*, Chapman & Hall.

[7] Dawid, A. P. (1994) "Selection Paradoxes of Bayesian Inference" in *Multivariate Analysis and its Applications* (Vol. 24), eds. T. W. Anderson, K. A.-T. A. Fang and I. Olkin, Philadelphia, PA: IMS.

[8] Dudoit, S., and Yang, Y.H. (2003) "Bioconductor R packages for exploratory analysis and normalization of cDNA microarray data" in G. Parmigiani, E. S. Garrett, R. A. Irizarry and S. L. Zeger, editors, *The Analysis of Gene Expression Data: Methods and Software*, Springer, New York. pp. 73-101.

[9] Efron B., Tibshirani, R., Storey, J. D., Tusher, V. (2001) "Empirical Bayes Analysis of a Microarray Experiment" *Journal of the American Statistical Association*, **96**, 1151-1160.

[10] Gelman A., Carlin J. B., Stern H. S., Rubin D. B. (2004) *Bayesian Data Analysis*, Chapman & Hall / CRC.

[11] Lohmueller K.E., Pearce C.L., Pike M., Lander E.S., Hirschhorn J.N. (2003), "Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease" *Nature Genetics* **33**. 177182.

[12] Mandel M. (2007) "Censoring and TruncationHighlighting the Differences" *The American Statistician*, **61**, 321-324.

[13] Mandel M., Rinott Y. (2009) "A Selection Bias Conflict and Frequentist Versus Bayesian Viewpoints" *The American Statistician*, **64**, 211-217.

[14] Scott J. G., Berger J. O. (2006) "An exploration of aspects of Bayesian multiple testing" *Journal of Statistical Planning and Inference*, **136** 2144-2162.

[15] Senn S. (2008) "A Note Concerning a Selection Paradox of Dawids" *The American Statistician*, **62**, 206-210.

[16] Smyth, G. K. (2005) "Limma: linear models for microarray data" in *Bioinformatics and Computational Biology Solutions using R and Bioconductor*, R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, W. Huber (eds.), Springer, New York, pages 397-420

[17] Soric B. (1989) "Statistical Discoveries and Effect-Size Estimation" *Journal of the American Statistical Association*, **84**, 608-610.

[18] Storey J. D. (2002) "A direct approach to false discovery rates" *Journal of the Royal Statistical Society: Series B*, **64** 479-498.

[19] Storey J. D., (2003) "The positive false discovery rate: A Bayesian interpretation and the q-value" *Annals of Statistics*, **31**, 2013-2035.

[1999] Yekutieli, D., Benjamini, Y.,(1999) "A resampling based False Discovery Rate controlling multiple test procedure" *J. Statist. Plann. Inference*, **82**, 171-196.

[20] Zhong H., Prentice R. L. (2008), "Bias-reduced estimators and confidence intervals for odds ratios in genome-wide association studies" *Biostatistics*, **9**(4):621-634.

[21] Zollner S, Pritchard J.K. (2007) "Overcoming the Winners Curse: Estimating PenetranceParameters from Case-Control Data" *The American Journal of Human Genetics*, **80**, 605 - 615.

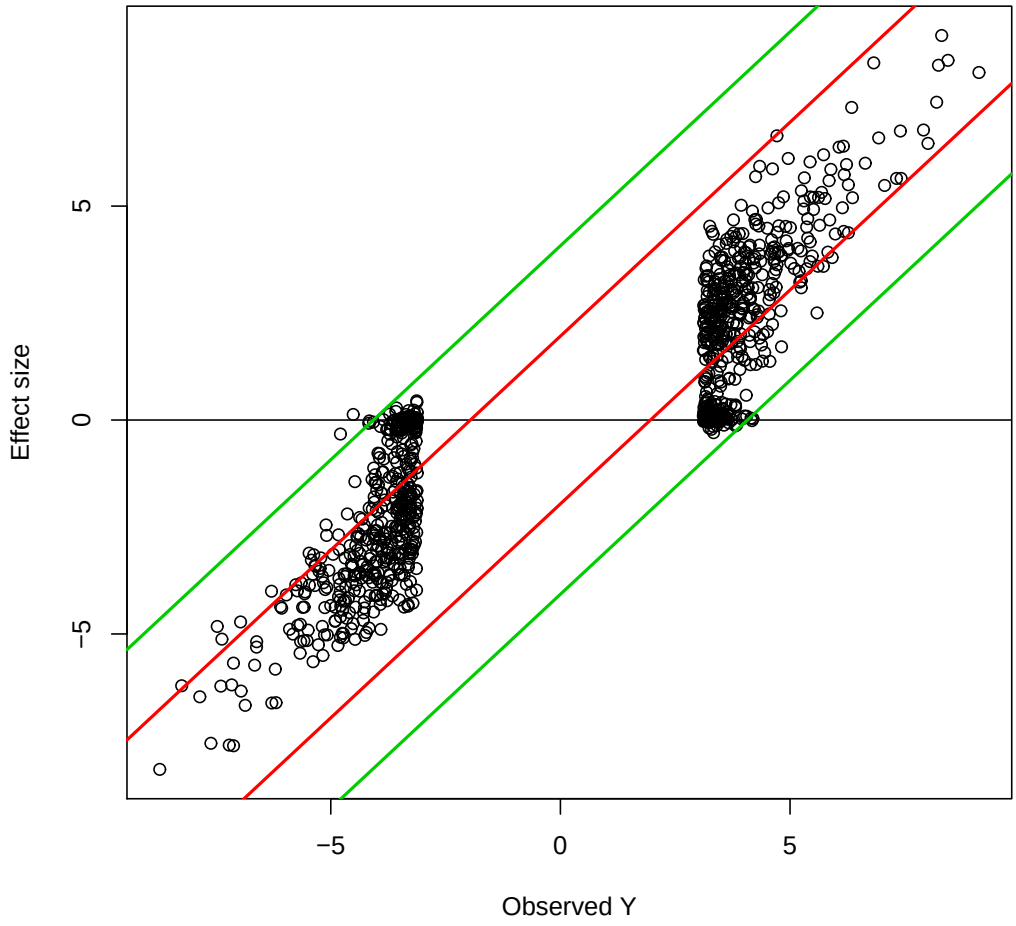

Figure 1: Simulated example – scatter plot of $|Y_i| > 3.111$ components. Y_i values are drawn on the abscissa of the plot, the ordinates are θ_i values. The red lines are marginal 0.95 CIs. The green lines are 0.05 FCR-adjusted CIs.

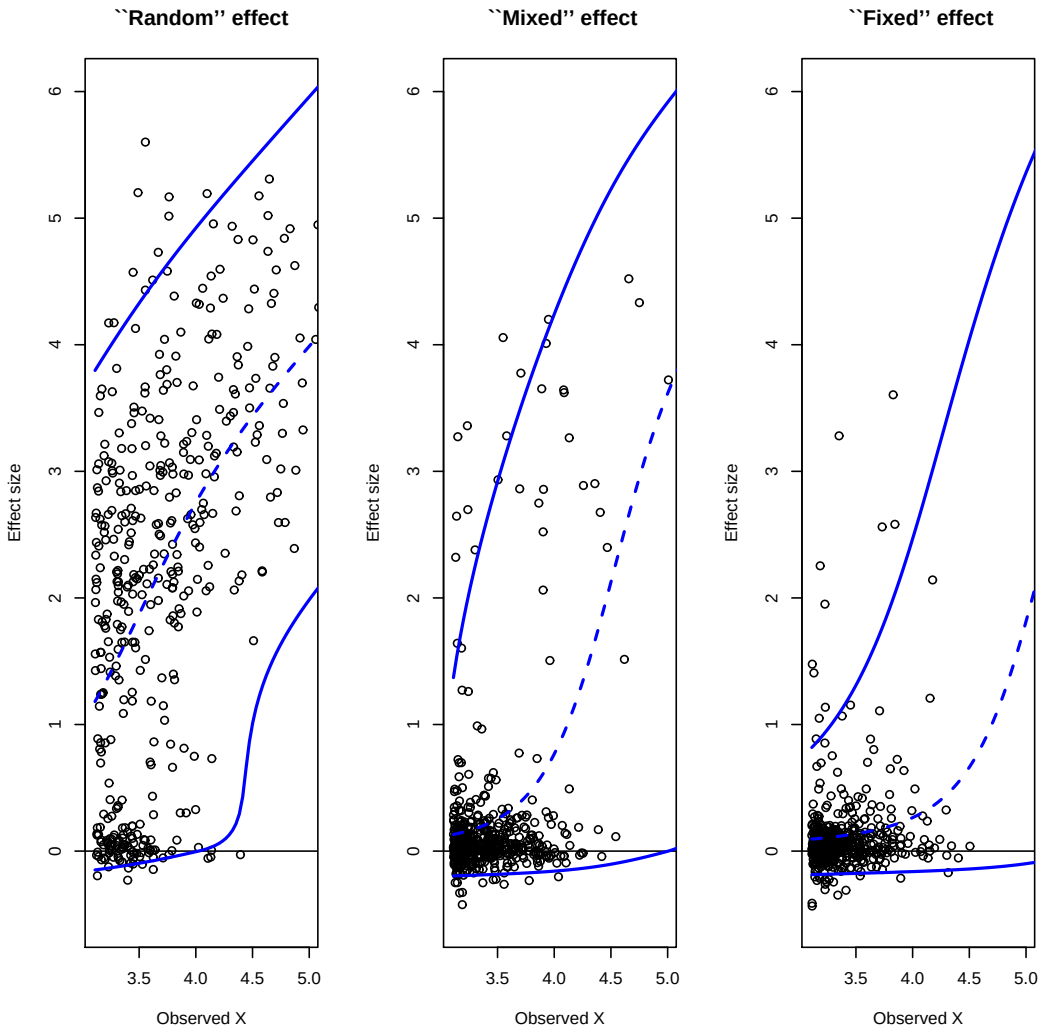

Figure 2: Simulated example – scatter plot of $Y_1 > 3.111$ realizations of (θ_1, Y_1) in the "random" effect truncated sampling model (left panel – 466 observations), the "mixed" effect truncated sampling model (middle panel – 498 observations), and the "fixed" effect truncated sampling model (right panel – 501 observations). The solid blue curves are the selection-adjusted 0.95 posterior credible intervals for θ_1, and the dashed blue curves are the selection-adjusted posterior means.

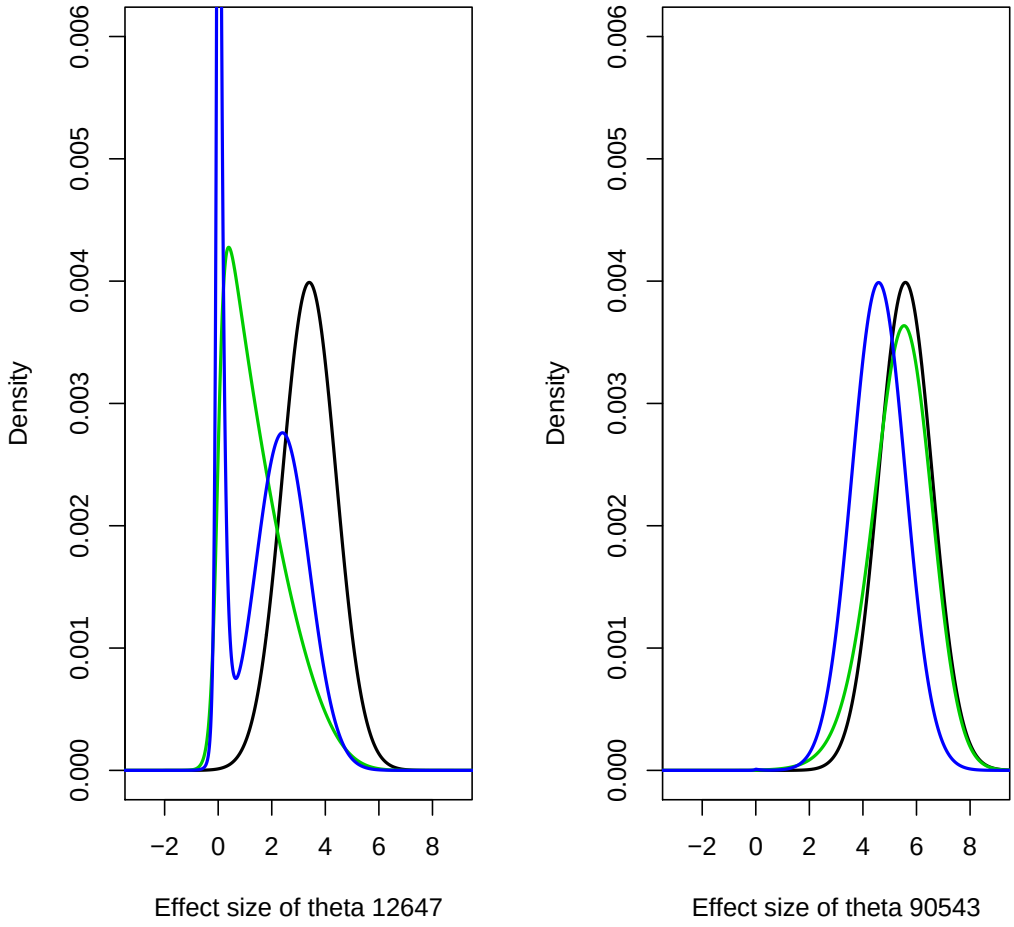

Figure 3: Simulated example – saBayes posterior distributions. The Posterior distributions for θ_{12647} are drawn in the left panel, the Posterior distributions for θ_{90543} are drawn in the right panel. The black curves are unadjusted posteriors; the blue curves are "random" effect model saBayes posteriors; the green curves are non-informative prior saBayes posteriors.

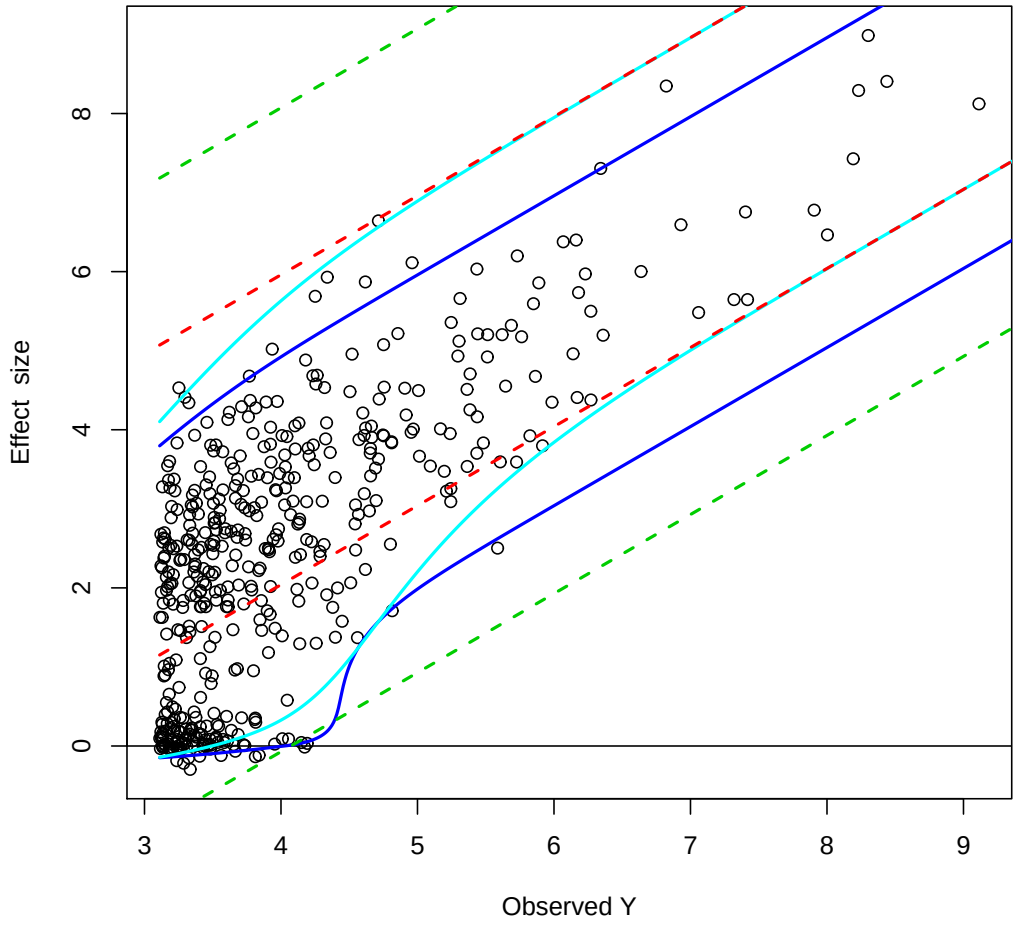

Figure 4: Simulated example – scatter plot of $Y_i > 3.111$ components. The dashed green and red lines are the CIs from Figure 1. The blue curves are the "random" effect model saBayes 0.95 credible intervals. The light-blue curves are the non-informative prior saBayes 0.95 credible intervals.

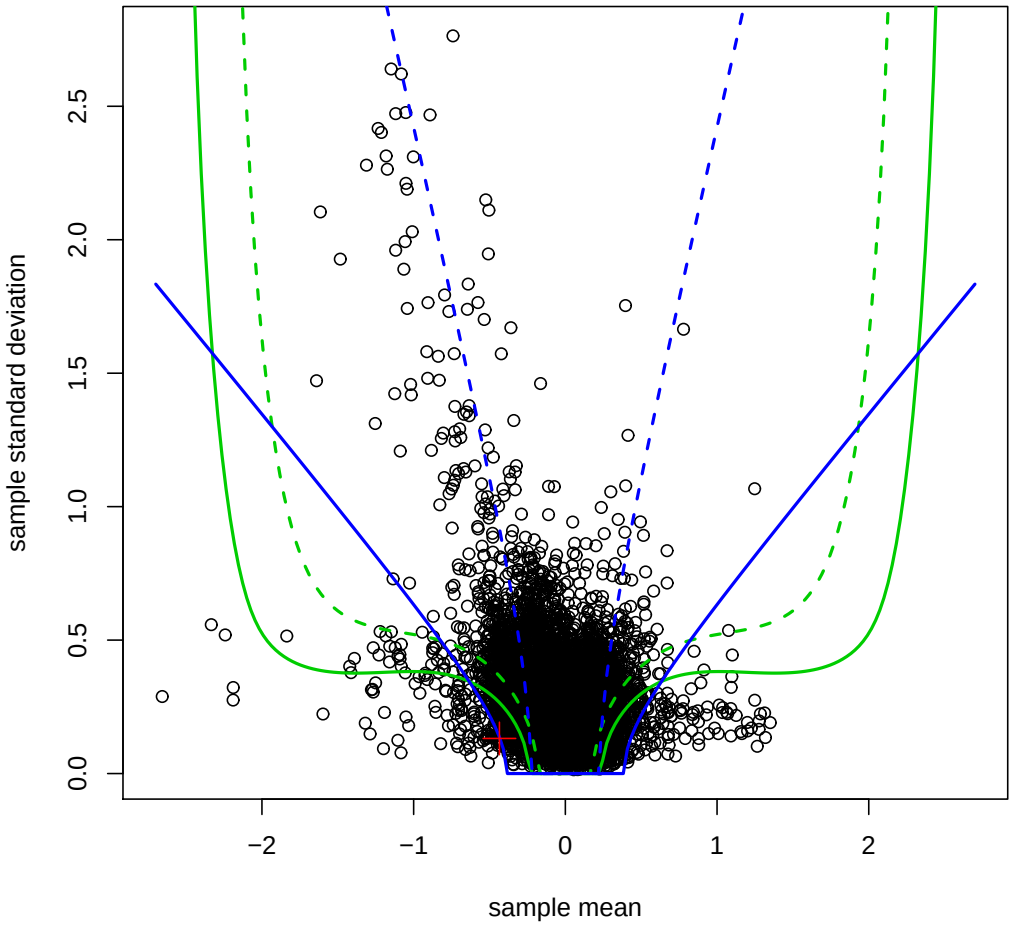

Figure 5: Swirl data – scatter plot of sample means and standard deviations. The abscissa of the plot is $\bar{y}_g$, the ordinates are s_g. The solid blue curve is $|\tilde{t}_g| = 4.479$. The dashed blue curve is $|\tilde{t}_g| = 2.64$. The solid green curve is $\tilde{\rho}(\bar{y}_g, s_g) = 0.05$. The dashed green curve is $\tilde{\rho}(\bar{y}_g, s_g) = 0.088$. The red plus sign is $(\bar{y}_{6239}, s_{6239})$.

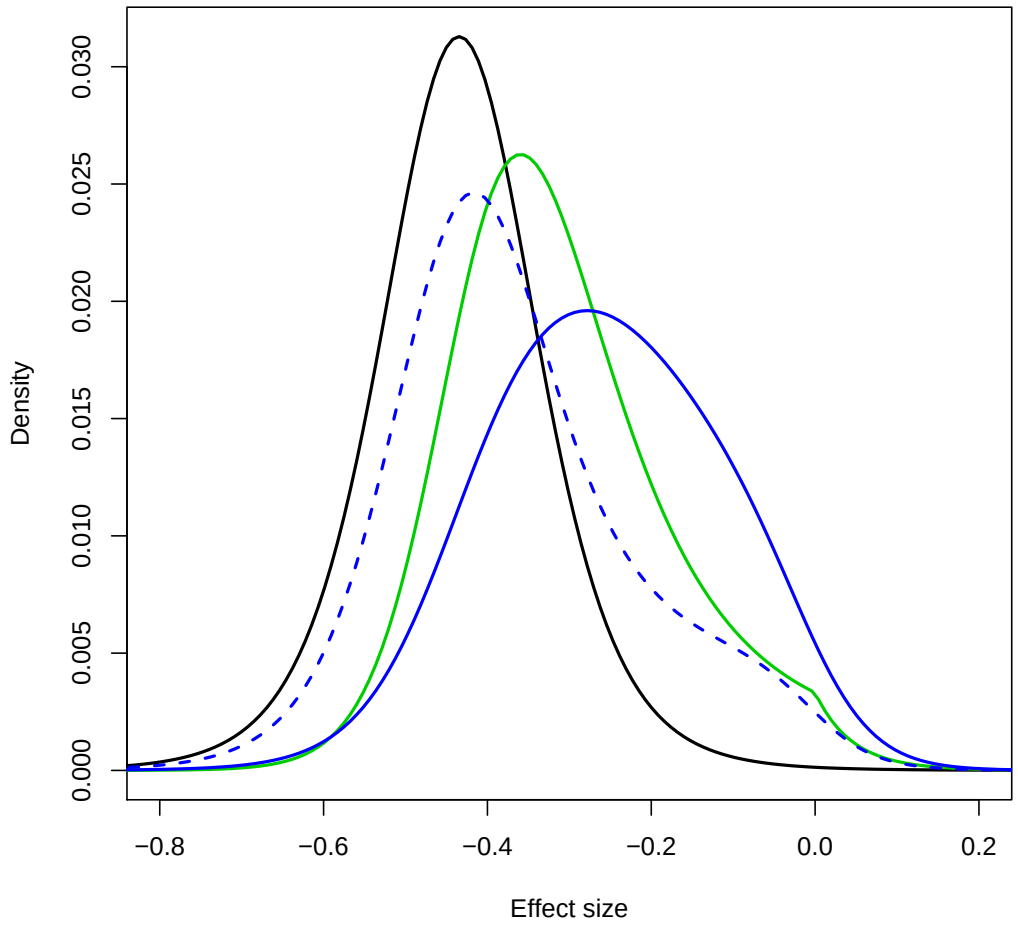

Figure 6: Swirl data – marginal posterior densities of μ_{6239}. The black curve is the non informative prior unadjusted posterior distribution. The green curve is the eBayes prior posterior distribution. The solid blue curve is the non-informative prior saBayes posterior distribution for the selection rule $|\tilde{t}_g| > 4.479$. The dashed blue is the non-informative prior saBayes posterior distribution for the selection rule$|\tilde{t}_g| > 2.64$.